

ANALIZA VALORILOR EXTREME

CRITERIUL LIMITEI DE TOLERANTA (CHAUVENET)

Suport teoretic

PRINCIPIUL: identificare unui **interval de probabilitate, centrat pe media unei distribuții normale**, care să includă **“aproape toate”** valorile prelucrabile.

Toate valorile plasate în afara acestui interval sunt considerate **extreme/suspecte** (X_s); cauzele apariției acestor valori sunt:

- variabilitatea selecției
- erorile de determinare.

Valoarea maximă a abaterii față de valoarea medie (D_{\max}) se alege în așa fel încât numărul de valori “eliminabile” (N_e) să nu fie prea mare ($N_e \leq 0,5$ pentru ca $N - 0,5 \approx N$), eșantionul rămas după filtrare să fie similar cu cel inițial:

$$N_e = N \cdot P\{|X - \bar{x}| \geq D_{\max}\} < \frac{1}{2} \Rightarrow P\{|X - \bar{x}| \geq D_{\max}\} < \frac{1}{2 \cdot N}$$

Criteriul de respingere este:

$$P\{X < \bar{x} - D_{\max}\} + P\{X > \bar{x} + D_{\max}\} = 2 \cdot \int_{X_m + D_{\max}}^{\infty} f_x = 2 \cdot P(Z) < \frac{1}{2 \cdot N} \approx 0$$

Probabilitatea valorilor eliminabile fără a fi afectată integritatea statistică a eșantionului, pe **fiecare** din cele **două** ramuri ale distribuției normale este:

$$P(Z) \leq \frac{1}{4 \cdot N}$$

Identificarea valorilor extreme se face pe baza ecuației

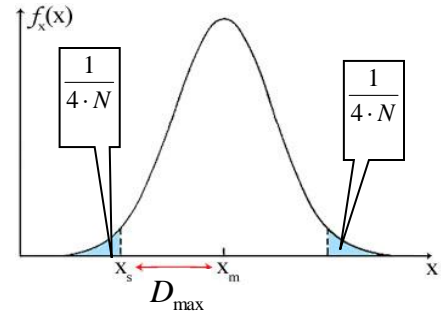
$$D_{\max} = \left| \frac{X_s - X_m}{s} \right| = Z = INV(P(Z))$$

în care

$$X_m = \frac{1}{N} \cdot \sum_{i=1}^{i=N} x_i \text{ -media aritmetică} \quad s = \sqrt{\frac{\sum_{i=1}^{i=N} (x_i - X_m)^2}{N - 1}} \text{ - abaterea standard}$$

$$x_{s_MINIM} \geq X_m - D_{\max} \cdot s \text{ limita inferioară a intervalului de toleranță}$$

$$x_{s_MAXIM} \leq X_m + D_{\max} \cdot s \text{ limita superioară a intervalului de toleranță}$$



Aplicație excel

Etapele de prelucrare într-o aplicație excel sunt:

- sortarea valorilor eșantionului de date
- diagram de variabilitate pentru identificarea calitativă a valorilor extreme
- calculul coeficientului de asimetrie
- reducerea gradului de asimetrie al selecției de date:

- $t = \sqrt{x}$

- $t = \ln(x)$

- $t = \frac{1}{x}$

- calculul parametrilor statistici descriptivi:

- media: $X_m = \frac{1}{N} \cdot \sum_{i=1}^{i=N} x_i$

- abaterea standard: $s = \sqrt{\frac{\sum_{i=1}^{i=N} (x_i - X_m)^2}{N - 1}}$

- calculul amplitudinii admisibile a selecției:

- $x_{extrem_MINIM} \geq \bar{x} - Z \cdot s$ cu $Z = NORMSINV\left(0.5 + \frac{2 \cdot N - 1}{4 \cdot N}\right)$

- $x_{extrem_MAXIM} \leq \bar{x} + Z \cdot s$ cu $Z = NORMSINV\left(0.5 + \frac{2 \cdot N - 1}{4 \cdot N}\right)$

- compararea valorilor considerate extreme cu limitele amplitudinii admisibile
- estimarea numărului optim de valori (N_{optim}) dintr-o selecție de N valori, corespunzător unui risc asumat (α) pentru intervalul de încredere corespunzător al valorii mediei de selecție ($\bar{x} - \mu$)

- $N_{optim} = \frac{t^2(\alpha, N - 1) \cdot s^2}{(\bar{x} - \mu)^2}$